Assessing Background Contaminant Concentrations

Presenter

Thomas Georgian, Ph.D.

Environmental and Munitions Center of Expertise, Environmental Sciences Division

US Army Engineering and Support Center, Huntsville

FY 16

U.S.ARMY

US Army Corps of Engineers
BUILDING STRONG®

Why is BG contamination important?

- Defense Environmental Restoration Program Manual (DoDM 4715.20) states:

  Risk assessments should not quantify exposure to naturally occurring substances present at concentrations unaffected by current or past site activity.

- EPA Policy Memorandum (May 2002):

  Generally, under CERCLA, cleanup levels are not set at concentrations below natural background levels. Similarly, for anthropogenic contaminant concentrations, the CERCLA program normally does not set cleanup levels below anthropogenic background concentrations…

Key Requirements for BG Studies

- **Hypothesis tests** to compare **mean** site and BG concentrations with quantitative tolerances for decision errors for Step 6 of the Data Quality Objective (DQO) Process.
  - UFP QAPP Worksheet #11 (Project/Data Quality Objectives)
- **Statistical sampling designs** for site and BG study areas for Step 7 of the DQO Process.
  - UFP QAPP Worksheet #17 (Sampling Design and Rationale)

Hypothesis Tests & Decision Errors

(Step 6)

Null hypothesis (H_0): $\mu_{Site} = \mu_{BG}$*

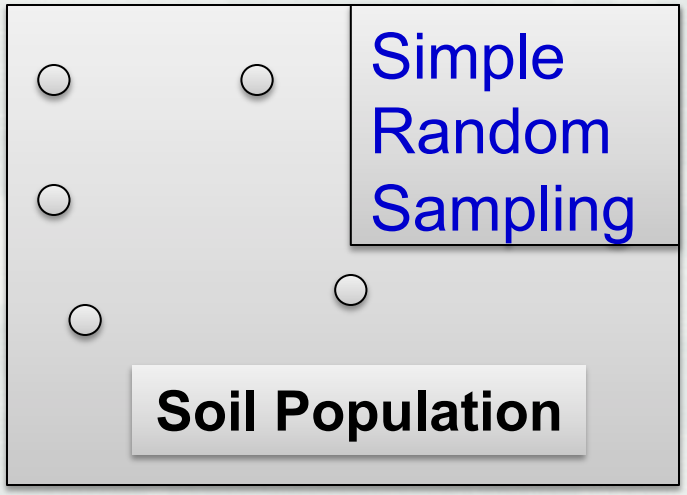
Alternative hypothesis (H_1): $\mu_{Site} > \mu_{BG}$

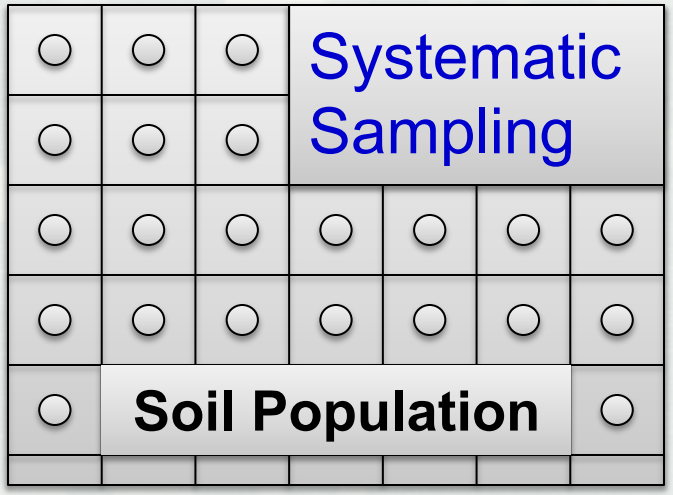
Site assumed "clean" until proven "dirty."

	H_0 True	H_0 False
Reject H_0	False Positive or Type I Error (α)	Correct Decision Power (1-β)
Do not reject H_0	Correct Decision Confidence Level (1-α)	False Negative or Type II Error (β)

* μ = "True" (population) mean

Statistical Sampling Designs
(Step 7)





- Reference: EPA QA/G-5S, *Guidance for Choosing a Sampling design for Environmental Data Collection*
- **Random** samples are needed for BG comparisons. A sample is random if every member of the population has an equal chance of being selected.

Number of Samples Needed

Visual Sampling Plan (VSP)

Input	→	Output
α, β, Δ/σ ↓		n_c ↑

σ = Standard deviation

Δ = $\mu_{Site} - \mu_{BG}$ = Difference important to detect to reject H_0 (e.g., $\Delta = 1 - 3\ \sigma$)

n_c = Minimum number of site & BG samples

Evaluating the Results

- ProUCL (Version 5) - Free EPA statistical software.

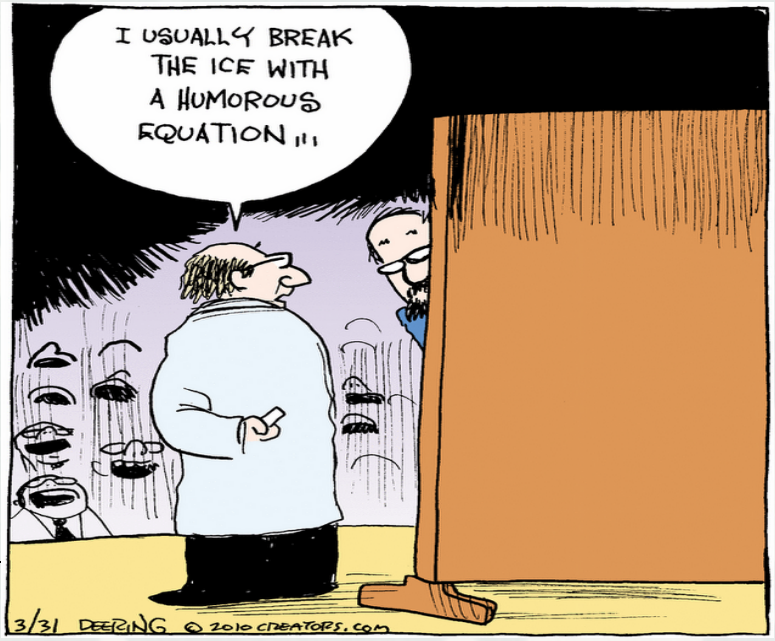
  http://www.epa.gov/osp/hstl/tsc/software.htm

- Calculates 95% UCLs of the mean using about 15 different methods & makes recommendations.

- Other capabilities: *Hypothesis tests* to compare mean site and BG concentrations; calculates various "not-to-exceed" ***background threshold values (BTVs)***.

Why is a statistical approach needed?

- Judgmental strategies often result in erroneous conclusions or decisions of unknown quality.
- Conclusions will be more scientifically defensible.
- Yields quantitative estimates of uncertainty.
- Mathematical equations with Greek symbols impress regulators & stakeholders.

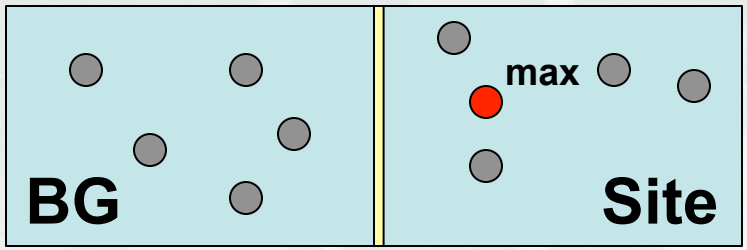


Common “judgmental” strategy for BG comparisons:

If site concentrations exceed the maximum background concentrations, it will be concluded site-related contamination is present.

- If the number of site samples (*k*) is equal to the number of BG samples (*n*), the false positive (FP) probability (wrongly concluding the site is “dirty”) α = 50%.

Flipping a coin gives the same FP probability!



Fun fact for people who love statistics

- The BG maximum is a type of BTV. The maximum of *n* BG results is a *(1- α)100% non-parametric k-k upper prediction limit (UPL).* The probability all *k* site results will be less this UPL (BG maximum) is:

$$1- \alpha = n/(n + k)$$

Note: When $n = k$, there is only a 50% probability all *k* site results will be less than the UPL ($1-\alpha = 0.5$).

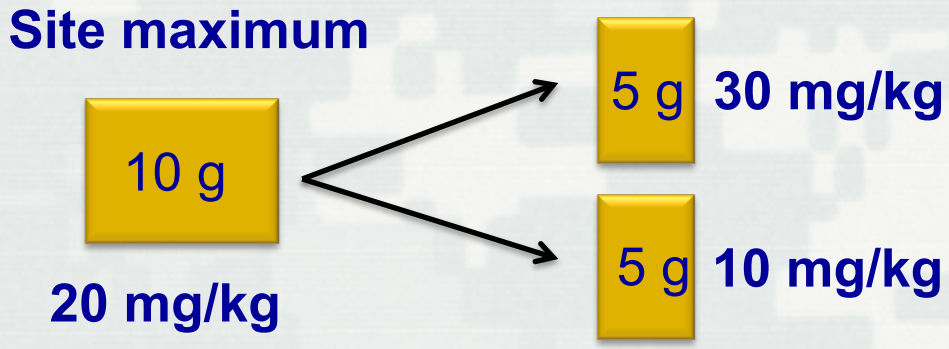
Decisions based on one data point

- Interstate Technology & Regulatory Council (ITRC) incremental sampling methodology (ISM) training (05/12):

  Be wary of decisions based on a single data point…
  The "maximum concentration" notion is meaningless...

- Owing to heterogeneity soil/sediment, results are highly variable and depend on the sample mass processed.

Example: BG maximum = 25 mg/kg.



More Invalid Comparisons Using Maxima

- Comparing maximum site concentration with BG 95% upper confidence limit (UCL) of the mean.
- Comparing maximum with an arbitrary multiple of the BG mean (e.g., 2 or 3 × BG mean).
- Comparing maximum with a BG 95% upper tolerance limit (UTL).
  - 5% of site results can exceed the BG 95% UTL when site and BG concentrations are not different. Determine whether > 5% of site results exceed the 95% UTL.

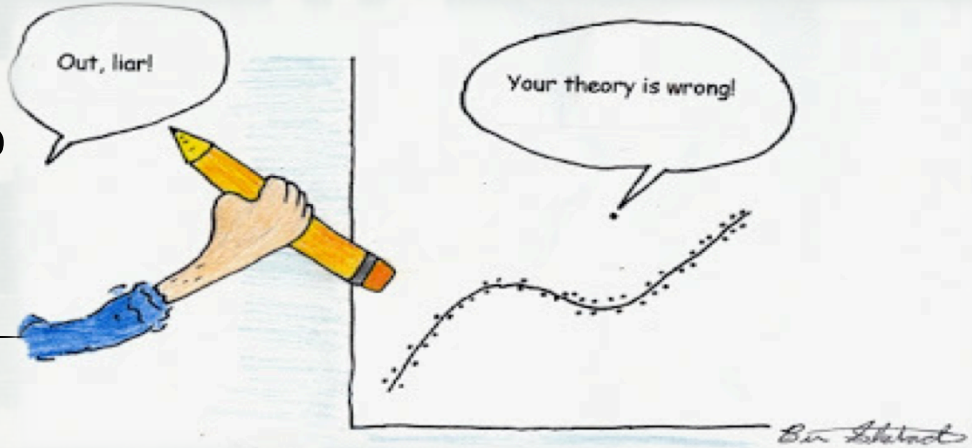
Background Threshold Values (BTVs)

EXPOSURE LIMIT 15

- If all site results are less than BTVs, it is assumed site and BG concentrations are consistent.
  - Examples: *Upper tolerance limits (UTLs)* & *prediction limits (UPLs)*.
- Inferior approach relative to hypothesis tests.
  - BTVs do not address false negatives; all site results below BTVs does not "prove" site is "clean."
  - Some site results can exceed a BTV when site and BGs means are not significantly different.
  - Use BTVs only when hypothesis tests cannot be done.

Common Practices that Result in FPs

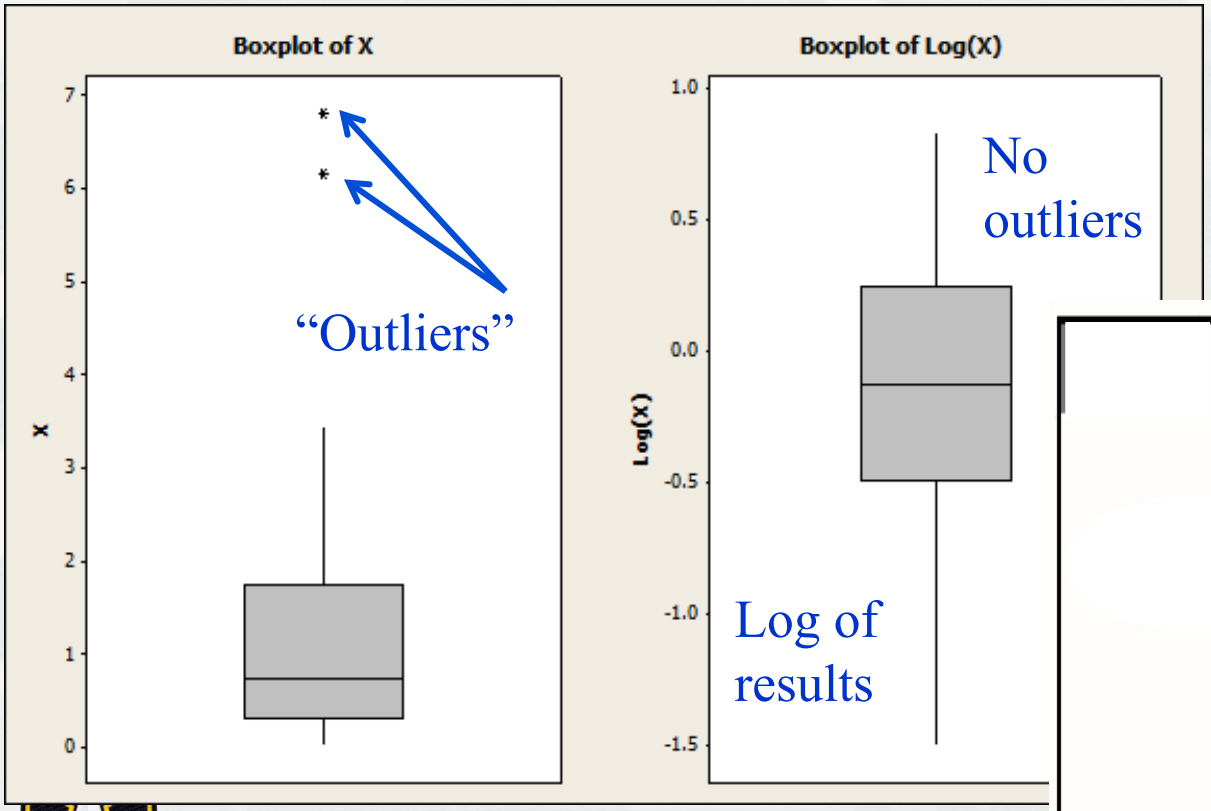
- Biasing BG samples to only “pristine” areas.
  - Anthropogenic BG contamination needs to be taken into account.
- Biasing site samples to only the most contaminated portions of the study area.
  - Statistical tests require representative (e.g., random) samples.
- Discarding BG data points on sole basis of statistical outlier tests (negatively biases BG data).
  - Physical justification for omitting BG points is also needed.



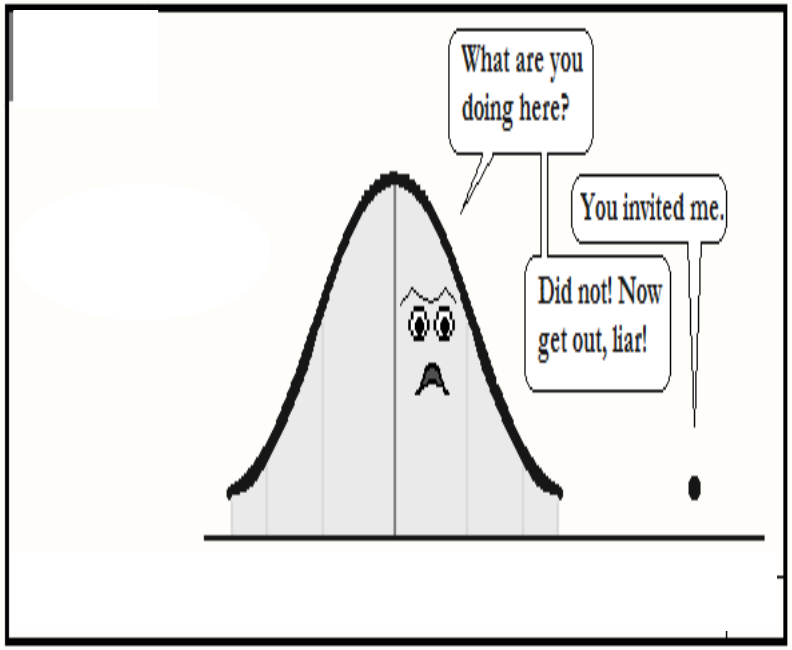
U.S.ARMY

Outliers (Continued)

- Outlier tests often assume normal distributions but there is no reason to assume BG data will follow a particular distribution. The figure below illustrates how box plots can misidentify outliers for non-normal data.

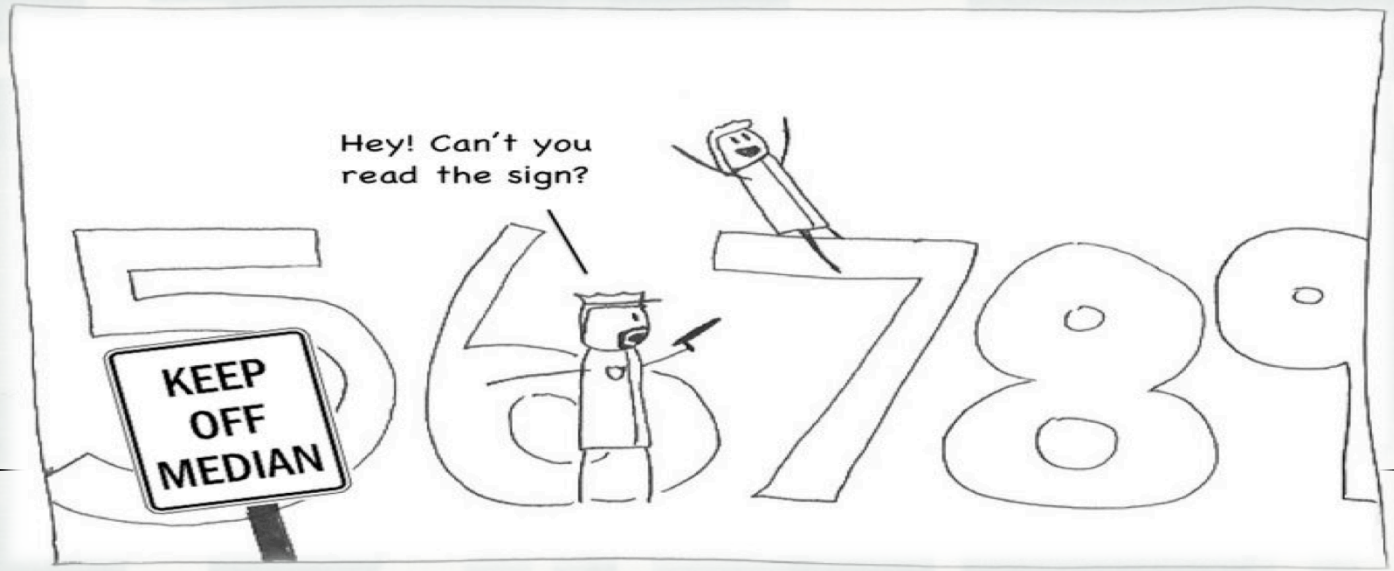


Box plots of same 20 points untransformed & log-transformed
from lognormal distribution.



Selection of Statistical Tests

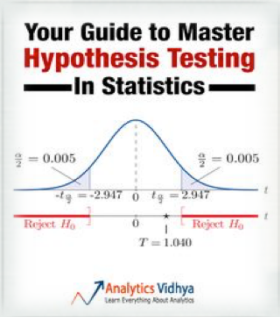
- Preferential use of hypothesis tests for comparing medians or other quantiles over tests for means.
  - Examples: Quantile, Slippage and Wilcoxon tests.
  - **Median ≠ Mean**. Use tests that compare means such as t-tests and permutation tests.



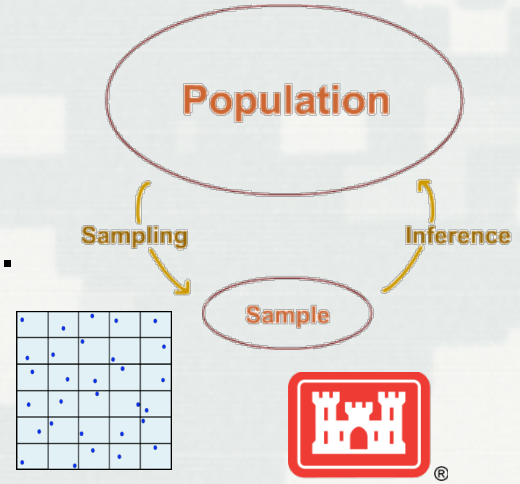
Conclusions

- Develop DQOs for BG evaluations using a statistical approach.

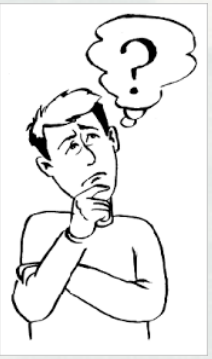
- Use hypothesis tests to compare site & BG **means** in Step 6 of the DQO Process.
  - ► Avoid comparisons with BTVs.



- Develop statistical sampling designs in Step 7 of the DQO Process.
  - ► Do conduct judgmental/biased sampling.



Questions?


?


NIH
GRANT
MACLEOD
"You are completely free to carry out whatever research you want, so long as you come to these conclusions."